Proceedings of the IASTED International
Conference
Intelligent Systems and Control 2000
August 14-16, 2000 – Honolulu, Hawaii,
USA

On-Chip Learning of Hyper-Spectral Data for Real Time Target Recognition

Tuan A Duong, Taher Daud, and Anilkumar Thakoor
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, CA 91109

Abstract: *It is well recognized that image based real time learning in neural computation using hyperspectral data set for target recognition is not only a complex problem to solve, but is also very time consuming. Solution of the same using neural processing with on-chip learning in hardware has never been attempted so far, even though such a high speed processing would be needed to perform real time data processing. It is important that a preprocessing step be included to cater to a high degree of input parallelism. Hence, a high-speed analog preprocessing method (developed and reported separately) that involved hardware-based data convolving scheme in a set of 3-D packaged chips with programmable templates, was used. Following that step, as the focus of our present paper, we have used the cascade error projection (CEP) learning algorithm (shown to be hardware-implementable) with on-chip learning (OCL) scheme to obtain three orders of magnitude speed-up in target recognition compared to software-based learning schemes. Thus, it is shown, real time learning as well as data processing for target recognition can be achieved. This paper first describes the processing details of a hyper-spectral data set in software containing a large 356,000 data points image. The paper also describes the fabrication of a test chip that was characterized and evaluated for comparison of its performance with the emulation results reaching up to 96.9% correct target recognition.*

I. Introduction:

Use of hyperspectral data for recognition of objects is known to offer advantages because of the differing reflective/ absorptive properties of the target materials involved over a range of frequencies in the visible and infrared regions. This expanse of wavelengths, however, brings about the complexity of massive amounts of radiation data to be sorted through. Using such data, artificial neural networks have been trained to recognize objects of interest [1].

At the same time, it is well recognized that learning in neural computation using hyperspectral data set for target recognition is not only a very complex problem to solve but is also a computationally intensive, and therefore, a time consuming process. Hence, real time solution of the same with on-chip learning in hardware has never been attempted so far.

By designing a suitable preprocessing step and combining it with a novel training scheme for the neural computation, an architecture has been developed that is particularly suitable for hardware implementation and more importantly, is capable of real time data processing.

Even though the emphasis of this paper is towards on-chip learning algorithm and its hardware implementation, we also briefly describe the preprocessing steps of the hyperspectral data set performed in software containing a large (356,000 data points) image.

This preprocessing step provides a compression of data into a manageable set of data, which are used for learning of the neural networks using a new

scheme. The data set is split into non-overlapping training, cross-validation, and testing data. The hardware implementation details are outlined and software and hardware results with different precision are compared and discussed.

II. Technical approach:

We have developed an on-chip learning (OCL) neural network technique that is capable of learning and adapting in real time. To do so, we have used the Cascade Error Projection (CEP) learning algorithm, which has been reported earlier [2-7]. CEP has been demonstrated as an efficient algorithm for hardware implementation due to its sufficiency with lower weight resolution and its quantization requirements. For example, it only required 4-bit or more weight quantization to learn 5-8 bit parity and chaotic time series prediction problems [2,5]. It is also faster learning algorithm since the learning is only conducted in the currently added hidden unit with respective weight changes.

Further, it is easy for the network to converge due to the cascading architecture, which avoids guessing and fixing a predetermined number of hidden units as is needed for Error Back-Propagation (EBP) learning technique.

Moreover, each added hidden unit is only parameterized by its set of weights that is correlated with the previous energy level. This fact allows us to manipulate the stepsize based on the previous energy and to exploit the full capability of the current set of weights [7,8]. In addition, CEP is theoretically proven for its convergence in Liapunov's sense when a cascading hidden unit is added [3,7,8].

We emphasize OCL architecture because it allows us to utilize the high speed of learning for real-time applications such as hyperspectral data processing for target recognition.

III. Architecture:

The architecture of CEP with OCL is shown in Figure 1. In this architecture, target and input signals are fed from outside, while output and conditioning signals are generated inside and brought outside. The output will provide actual performance of the network and conditioning signal is used to monitor the status of learning system.

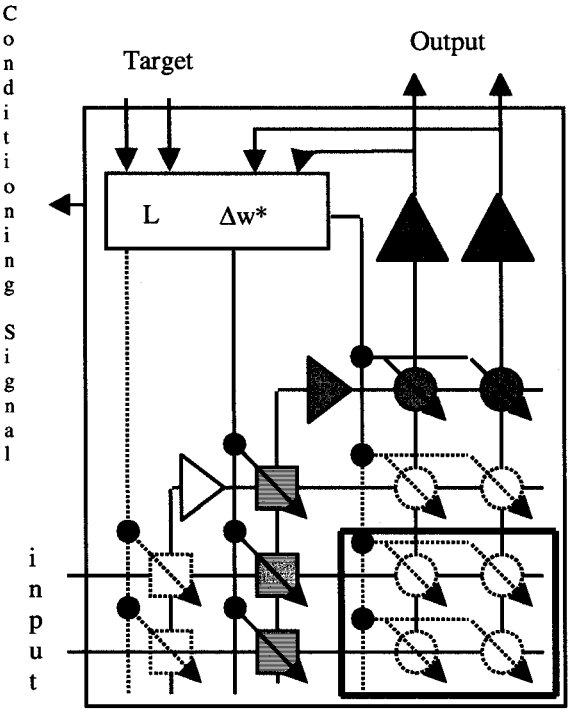

Figure 1: An OCL architecture is shown where Δw* is generated and updated on-chip. Blank squares and circles denote weight components that are trained and frozen, while shaded squares and circles are undergoing learning.

Initially, input is directly mapped to output by set of weights in the rectangular box. This weight set is obtained by perceptron learning. Then, a hidden neuron is introduced and the input and target are fed into neural network and Δw^* is generated from block A (a part of L), which go to update the weights represented by shaded squares and circles until the network satisfies user's criteria. If not, iteratively hidden neurons are introduced and weight updates obtained.

To make the algorithm hardware-implementable, we break block L into three blocks: A, B, and C. Block A provides $\frac{\partial \Phi}{\partial S}$, block B provides the updating weight $\frac{\partial \Phi}{\partial S}\frac{\partial S}{\partial W}$ for the shaded squares, and block C calculates the output weights for each added hidden unit.

- Block A:

$$(\frac{\partial \Phi}{\partial S})^p = f_h^{\prime p}(n+1) - \frac{1}{m}\sum_{o=1}^{m}(t_o^p - o_o^p))f_h^{\prime p}$$

- Block B:

$$\Delta w_{ih} = -\eta \mathrm{I}^p(n)(\frac{\partial \Phi}{\partial S})^p$$

- Block C:

$$w_{ho}(n) = \frac{\sum_{p=1}^{P}(t_o^p - o_o^p)f_o^{\prime p} f_h^p(n+1)}{\sum_{p=1}^{P}\{f_o^{\prime p} f_h^p(n+1)\}^2}$$

where

$$I(n) = \begin{bmatrix} X_i \\ X_h(n) \end{bmatrix}$$

$$S = \mathrm{I}^T(n) W_{ih}(n+1)$$

$$\Phi^p = \{f_h^p(n+1) - \frac{1}{m}\sum_{o=1}^{m}(t_o^p - o_o^p)f_o^{\prime p}\}^2$$

where n is the number of added hidden units, p, number of training patterns, and m, number of output units. Further, o denotes output index, h is an index of hidden units, and η is the learning rate. The index of weight from input to hidden unit is denoted by *ih*, and X is the input vector.

For each learning iteration, energy is monitored and accumulated for evaluation of the condition of learning system. When energy accumulation for all input pattern is above a threshold level, the learning is continuing on. Otherwise, it is terminated.

IV. On-Chip Learning:

In CEP architecture, important features of OCL consist of *independent layer learning* and *dynamical stepsize manipulation.*

- *Independent layer learning*:

The current error surface that did not meet the stopping criteria for learning as provided by user is projected to a newly added hidden unit for learning. This learning is only conducted *independently* within a new weight set (the previous weight sets are frozen) (See Figure 1). Because of independent layer learning, the learning weight set can be set a stepsize as a gain factor of synapse. The manipulation of stepsize in learning phase is shown in the previous report that it helps to reduce weight quantization requirement [7-8].

- *Dynamical stepsize:*

Based on our previous study, the stepsize of the current energy is roughly proportional to the previous energy level [3,7]. Independent layer learning allows setting a stepsize as needed to enhance the learning capability under bit quantization constraints.

The simulation results were shown to be encouraging in previous studies by using dynamical stepsize e.g., 5-8 bit parities, chaotic time series prediction, and color segmentation problems [2,6,9].

V. Learning with limited weight quantization:

One of the most challenging parts for OCL is the convergence problem. There are several factors that can cause learning divergence. In case of hardware, the two obvious reasons are the architecture and the availability of the weight quantization.

- *Architecture:* This is one of the most crucial factors. Fixed and predetermined architecture is a poor choice since the learning convergence is dependent on training data structure, learning rate, and initial random weight. This architecture is a flexible quick fix, but it is not optimal architecture since learning is only conducted in the sub-weight space. In the CEP with OCL, we have adopted the cascading architecture.

- *Weight quantization:* This step is very expensive in VLSI hardware implementation. Increasing even one bit in synaptic weights requires almost double the space on silicon, and twice the power consumption [10]. In addition, synaptic weight requires very large population in a system. Further improvement in quantization may not be possible when synaptic weights reach to certain bit quantization level. In our study, we limited ourselves to 8-bit weight quantization which had been designed and tested [10-11].

VI. Application:

The Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) is an optical sensor that delivers calibrated images of the upwelling radiance in 224 spectral channels, or bands, with wavelengths from 400-2500nm. AVIRIS collects spectra sequentially by using a Whisk-broom scan mechanism. The radiance from an approximately 20 m^2 patch on the ground is dispersed thru four grating spectrometers to obtain a spectrum consisting of 224 channels [12]. The AVIRIS data is inverted to units of spectral reflectance using a radiative transfer model estimate of atmospheric path radiance and reflected radiance [12]. This data were played as background scene, while target spectra were obtained from ground truth measurements. In these experiments 10% of the target data were mixed into AVIRIS scene. The details of this work are published elsewhere [13].

From the mixing background and target data, the composite data is highly redundant and noisy. Directed Principle Component Analysis (DPCA) is used to reduce the data dimension and to get rid of white noise from it [13]. The original pixel dimensions are reduced from 224 bands to 16 bands without much loss of information.

In 365,000 data points, 2,224 data points were selected of which each pixel consisted of 16 elements for training. Similarly, another 2,224 non-overlapped data points were selected for cross-validation which allowed examination of the status of learning in order to avoid over-learning problem. Finally, 17,821 data set that was again non-overlapping with the training and cross-validating data sets was used to test and evaluate the system.

VII. Simulation results:

In this simulation, we have studied bit constraints under learning neural network. We started with floating point machine (64-bit with double precision), and later reduced it to 8- through 5-bit weight quantization. The results are shown in Table A.

Table A: H/W emulation and S/W results

Weight quanti-zation	Training	Cross validation	Testing results
64-bit	99.9%	99.9%	99.9%
8-bit	97.4%	97.2%	96.9%
7-bit	94.4%	95.1%	94.2%
6-bit	88.6%	90.9%	87.8%
5-bit	86.0%	87.1%	85.7%

Each hidden unit was required to perform 1000 iterations to complete learning for that hidden unit. Using SPARC ULTRA2, it required 20 min to learn 2,224 input patterns. To project this learning in time consumption, each iteration was designed to take 0.25μs and it consumed 3.36 seconds to complete the learning. In addition, digital computer would require more time to train when dimensionality would increase. On the other hand, OCL training time is constant regardless of the dimensionality of the problem. From this comparison, we were able to ascertain that about 3 order of magnitude faster training was obtained using OCL approach.

VIII. Conclusion:

In this paper, we demonstrate that OCL using CEP architecture would require 7-bit weight or more to obtain above 94% correct recognition using sub-pixel hyperspectral data. This learning is an improvement of three orders of magnitude in training time compared to that in simulation.

Even simulation with 8-bit weight quantization achieved 96.9% recognition, which is a first ever. Downloading of the simulation-derived weights into the test chip with 7-bit weight quantization and subsequent hardware in the loop learning resulted in the efficiency of target recognition of 88%. With 20% target mix, this is the first ever hardware result for the solution of a hyperspectral target recognition data processing based on on-chip learning. It is expected that with further hardware parameter modifications, a substantial closeness to simulated results will be achieved.

Acknowledgments:

The research described herein was performed by the Center for Space Microelectronics Technology, Jet Propulsion Laboratory, California Institute of Technology and was jointly sponsored by the Ballistic Missile Defense Organization (BMDO), and the National Aeronautics and Space Administration (NASA). The authors would like to thank D. Weldon, and Dr. C. Padgett for providing data and preprocessing them.

References:

1. A. Howard, C. Padgett, "*Generalized approach to real time pattern recognition of Science data*," Pattern Recognition, 32, 2069-2071, 1999.
2. T.A. Duong, "*Cascade Error Projection-an efficient hardware learning algorithm*," Proceeding Int'l IEEE/ICNN in Perth, Western Australia, Oct. 27-Dec 1, 1995, vol. 1, pp. 175-178.
3. T. A. Duong, *Cascade Error Projection-an efficient hardware learning theory, Ph.D. Thesis, UCI*, 1995.
4. T. A. Duong, et al., 1996, "Cascade Error Projection-A New Learning Algorithm," Proceeding Int'l IEEE/ICNN in Washington D.C., vol. 1, pp. 229-234.
5. T.A. Duong and T. Daud, "*Cascade Error Projection_a Learning algorithm for hardware implementation*," Lecture Notes in Computer Science-Foundation and Tools for Neural Modeling, vol. 1, pp450-457, IWANN'99, Alicante, Spain, June 1999.

6. T.A. Duong and T. Daud, "*Cascade Error Projection with low bit weight quantization for high order correlation data*", paper 2144, session 5-8, INNS/NNC of IEEE in Washington D.C., 1999.
7. T. A. Duong and A. R. Stubberud "*Convergence analysis of cascade error projection-an efficient learning algorithm for hardware implementation*", To be publish in International Journal of Neural System, 2000.
8. T.A. Duong, "*Train Neural networks with fewer quantization bits,*", *NASA Tech nology Briefs*, vol. 23, No. 7, pp. 56-57, July 98.
9. E. Fiesler, S. Campbell, L. Kempen, and T. Duong."*Color Sensor and Neural Processor on One Chip.*" International Symposium on Optical Science, Engineering, and Instrumentation, Proc. of the SPIE, vol. 3455 'Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation', pp. 214-221, 1998.
10. T.A. Duong, S. Kemeny, T. Daud, A. Thakoor, C. Saunders, and J. Carson, "*Analog 3-D Neuroprocessor for Fast Frame Focal Plane Image Processing*," *The Industrial Electronics Handbook*, Chap. 73, Ed.-In-Chief J. David Irwin, CRC PRESS, 1997.
11. T. A. Duong et. al, "Low Power Analog Neurosynapse Chips for a 3-D "Sugarcube" Neuroprocessor," *Proc. of IEEE Intl' Conf. on Neural Networks* (ICNN/WCCI, June 28-July 2, 1994, Orlando, Florida), Vol III, pp. 1907-1911.
12. G. Vane, et al. "The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)", *Remote Sens. Environ.* Vol. 44, pp. 127-143, 1993.
13. A. Howard, et al., "Intelligent Target Detection and Hyperspectral Imaging", Proc. 13th International Conf. Applied Geological Remote Sensing, Vancouver, British Columbia, Canada, March, 1999.